

AD-A103 023

BIOTECHNOLOGY CONFERENCE: PROTEIN ENGINEERING HELD IN
OXFORD UNITED KINGDOM ON 5-8 APRIL 1987(U) OFFICE OF
NAVAL RESEARCH LONDON (ENGLAND) C E ZOMZELY-NEURATH
27 JUL 87 ONRL-7-019-C

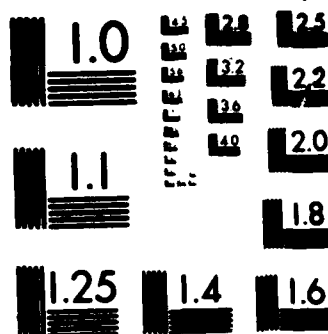
1/1

UNCLASSIFIED

F/G 6/1

ML

END
9-87
DTIC



MICROCOPY RESOLUTION TEST CHART
 NATIONAL BUREAU OF STANDARDS-1963-A



DTIC FILE 100/1

ONRL Report 7-019-C

AD-A183 823

DTIC
ELECTE
AUG 24 1987
S D

Biotechnology Conference: Protein Engineering '87
University of Oxford, UK

Claire E. Zomzely-Neurath

27 July 1987

Approved for public release; distribution unlimited

U.S. Office of Naval Research, London

87 8 17 088

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A183 823

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		4. PERFORMING ORGANIZATION REPORT NUMBER(S) 7-019-C	
5. MONITORING ORGANIZATION REPORT NUMBER(S)		6a. NAME OF PERFORMING ORGANIZATION US Office of Naval Research Branch Office, London	
6b. OFFICE SYMBOL (If applicable) ONRL		7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code) Box 39 FPO, NY 09510		7b. ADDRESS (City, State, and ZIP Code)	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code)		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Biotechnology Conference: Protein Engineering '87, University of Oxford, UK			
12. PERSONAL AUTHOR(S) Claire E. Zomzely-Neurath			
13a. TYPE OF REPORT Conference	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) 23 July 1987	15. PAGE COUNT 11
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD 06	GROUP 01, 02	Proteins Amino acids Biotechnology; Mutagenesis Synthesis	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>Presentations given at this major international conference held at the University of Oxford, UK, in April 1987 are reviewed. The topics include theoretical aspects of protein structure, protein structure and dynamics, protein engineering methods, protein engineering stability, protein engineering binding and catalysis, and protein engineering medical and industrial applications.</p>			
DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
NAME OF RESPONSIBLE INDIVIDUAL J. J. FOX		22a. TELEPHONE (Include Area Code) (44-1) 409-4340	22c. OFFICE SYMBOL 11

ORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

CONTENTS

	<u>Page</u>
1 INTRODUCTION	1
2 THEORETICAL ASPECTS OF PROTEIN STRUCTURE	1
3 PROTEIN STRUCTURE, STABILITY, AND DYNAMICS	3
4 PROTEIN ENGINEERING: METHODS	5
5 PROTEIN ENGINEERING: BINDING AND CATALYSIS	6
6 PROTEIN ENGINEERING: MEDICAL AND INDUSTRIAL APPLICATIONS	9
7 CONCLUSION	11

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	



BIOTECHNOLOGY CONFERENCE: PROTEIN ENGINEERING '87, UNIVERSITY OF OXFORD, UK

1 INTRODUCTION

This major international conference took place at the University of Oxford, UK, from 5 through 8 April 1987. About 450 scientists from 12 European countries as well as the UK, Israel, Japan, US, Canada, Australia, and Eastern European countries (USSR, Hungary) attended the conference. Scientists from the UK constituted 33 percent of the total delegates with France and West Germany having the highest representation among the European countries. About 30 percent of the attendees represented industrial organizations with the balance from academic institutions.

The meeting was organized by A.R. Rees (Laboratory of Molecular Biophysics, University of Oxford) and G.A. Petsko (Department of Chemistry, Massachusetts Institute of Technology, Cambridge), who are the executive editors of the new journal, *Protein Engineering*, published by IRL Press (Oxford/Washington, D.C.).

The conference was divided into six symposium sessions on the following topics:

- Theoretical aspects of protein structure
- Protein structure and dynamics
- Protein engineering--methods
- Protein engineering--stability
- Protein engineering--binding and catalysis
- Protein engineering--medical and industrial applications.

There were also poster sessions, comprising 107 posters, as well as a trade exhibition. Protein engineering is characterized by the coordinated application of specialized theoretical and experimental disciplines. These include the determination and prediction of protein structure, experimental studies of protein folding, physical and biochemical methods for correlating protein structure and activity, chemical modification, and

generalized and site-specific mutagenesis. Among the protein engineer's objectives are the investigation of those principles which govern structure, the determination of the mechanisms through which particular functions are expressed, and the empirical testing of these principles by introducing site-specific alterations and then evaluating any changes in structural or functional properties. Ultimately, protein engineering has the potential to be a source of novel proteins with novel applications in technology, medicine, and industry.

A great deal of interesting and informative material was presented at this conference at Oxford. Thus, summaries of only a few selected presentations are given in this report.

Unfortunately, no proceedings will be published, according to A. R. Rees, organizer of the meeting. Abstracts of the meeting will, however, be published shortly in an issue of the journal, *Protein Engineering*.

2 THEORETICAL ASPECTS OF PROTEIN STRUCTURE

The topic of protein folds and protein sequences was discussed by C. Chothia (MRC Laboratory of Molecular Biology, Cambridge, UK). He emphasized that the successful prediction of the structure of a protein from its sequence, using the known structure of an homologous protein is of great importance for protein engineering. It is also important for testing theories or models for the sequence-dependent aspects of protein conformation. Chothia reviewed the results that he and his group have obtained in the cases where (1) the proteins have similar folds but very different sequences, (2) the proteins have low to moderate sequence identities, and (3) the proteins have high sequence identities. He stated that, essentially, the major response of proteins to mutation, whether by natural evolution or experimental mutagenesis is a conformational change and that functional restraints determine which of the conformational changes are accepted.

Modeling of protein structure using data bases was discussed by M.J.E. Sternberg (Department of Crystallography, Birbeck College, London, UK). The known protein structures have been organized in a relational data base that stores the Brookhaven coordinates and computer-derived conformational features such as secondary structure, dihedral angles, and solvent accessibility. The relational data base enables rapid interrogation of features. Applications include the selection of loop conformations for predicting unknown structures based on the x-ray coordinates of an homologous molecule and also for obtaining conformational features (i.e., disulfide bridges) for modeling the consequences of site-specific mutations. The protein sequence data base can be used to obtain a multiple alignment of homologous sequences. The combined information from all the sequences can be used to improve secondary structure prediction by 10 percent and to locate probable active site residues in protein families.

A report on the structure, design, and modification of loop regions in proteins was presented by J.M. Thornton (Birbeck College, University of London, UK). In proteins the loop regions connecting the secondary structures comprise about 30 percent of the structure. The loop regions are on the surface of the protein, and are often flexible. Since insertions and deletions in homologous sequences usually occur in these regions, it is expected that engineered mutations will be most easily tolerated in the loop regions without destroying the three-dimensional structure, according to Thornton. Detailed analysis of the conformations of loop regions in the proteins of known structure by Thornton and her group has revealed that for short loops there are structural families with specific sequence patterns. Such patterns can be used to guide the choice of sequences in novel protein design or in site-directed mutagenesis experiments. Loops are widely involved in recognition between molecules, such as antibody-antigen interactions and protein-receptor recognition. Peptides, excised from loop regions, can

be used to elicit an immune response against the native protein.

The correlation of coordinated amino acid substitutions with function in tobacco mosaic viruses (TMV) was discussed by D. Altschuh (MRC Laboratory of Molecular Biology, Cambridge, UK). TMV is the best studied example of a self-aggregating system. Sequence data are available for the coat proteins of six TMV's, with homologies ranging from 33 to 82 percent; atomic coordinates are known for TMV wild-type. According to Altschuh, the constraints on the overall size and shape of the protein subunit and on the character of those regions of the subunit surface involved in quaternary structures should be reflected in the nature and pattern of acceptable amino acid substitutions. She and her group have found a significant spatial relationship between groups of residues with identical amino acid substitution patterns. The relationship strongly suggests, according to her, that after mutation, the groups of residues have not become stabilized independently of each other, and that their location is linked to a particular function--at least in viruses identical to the disulphide in these residues. The most conserved feature of TMV is the ribonucleic acid (RNA) binding region. Altschuh and her group found that core residues are conserved in all viruses or show mutations complementary in volume. The specificity of inter-subunit contacts is achieved in different ways in the three more distantly related viruses. The strategy described above for detecting coordinated substitutions has worked well within the TMV family where the protein has extensive quaternary structure, according to Altschuh. If this approach can be applied equally successfully to other families of proteins, it could contribute to the understanding of protein folding and interactions.

The topic of messenger RNA (mRNA) translation and protein folding *in vivo* was discussed by J.C. Swaffield (Biotechnology Unit, Institute of Genetics, University of Glasgow, UK). Many highly expressed genes from *Saccharomyces cerevisiae* show a strong bias in their choice

of codons for the 20 amino acids. This codon bias correlates strongly with the relative abundance of the iso-accepting transfer RNA's (tRNA). According to Swaffield the set of preferred codons varies between organisms. When "rarely used" codons are clustered within a gene, a pause in the rate of translation is predicted that would result in the accumulation of nascent polypeptide chains of a discrete length. Since proteins are synthesized from the N-terminus and initial folding reactions probably occur before translation is completed, Swaffield believes that translational pauses possibly influence the folding of some proteins by allowing regions of the growing polypeptide chain to fold correctly before C-terminal regions are synthesized. According to Swaffield, a major problem in relating potential translational pauses to protein folding is the lack of knowledge about the tertiary/quaternary structures of many proteins for which the gene sequence is available. However, there is a tight correlation between the presence of potential translational pauses and the interdomain regions of the *arom* multifunctional enzyme in *Saccharomyces cerevisiae*. According to Swaffield, potential translational pauses have been observed in the genes for other multifunctional enzymes in yeast (e.g., TRP3, TRP5), but in these cases the pauses do not seem to lie in interdomain regions.

3 PROTEIN STRUCTURE, STABILITY, AND DYNAMICS

Proteins are well-designed for their functions. They may be rigid or flexible to various degrees as required for optimal performance. Flexibility at the level of amino acid side-chains occurs universally and may be important for some functions. Large-scale flexibility where large parts of a protein rearrange or move coherently is particularly interesting and was discussed by R. Huber (Max Planck Institute for Biochemistry, Martinsried [Munich], West Germany). Huber and his group have carried out extensive

research on the flexibility and rigidity requirements for functions of proteins and protein-pigment complexes.

Huber and his group have found that it is possible to differentiate between different categories of flexibility, order-disorder transitions of domains, and domain motions. The domains may be flexibly linked to allow rather unrestricted motion or the motion may be constrained to certain modes by hinges. The connecting segments and the hinges show characteristic structural features. The following examples presented by Huber illustrate various aspects. Small proteinase inhibitors are essentially rigid molecules and provide tight complementary binding to their cognate protease. The large plasma inhibitors, however, exhibit large conformational changes upon interaction with proteases, probably for regulatory purposes.

Huber and his group found that the pancreatic serine proteases exhibit a disorder-order transition of their active domain between proenzyme and enzyme forms as a means of regulating enzymic activity. Immunoglobulins show rather unrestricted and also hinged domain motions in different parts of the molecule, probably to allow cross-linking of antigens. Citrate synthase adopts open and closed forms by a hinged domain motion to bind substrates and release products and to perform the catalytic condensation reaction respectively, according to Huber. In the multienzyme complex riboflavin synthase in which two consecutive enzymic reactions are catalyzed by two distinct enzymes it was found that restricted motions by engaging one enzyme in a capsid formed by the other, as well as directed substrate delivery, seem to play an important role in enzymatic activity. In contrast to the above examples, Huber stated that motion would be deleterious to function in the light-harvesting complexes and the reaction centers involved in the photosynthetic light reactions. These complexes are huge protein complexes which serve as matrixes to hold the pigments active in light absorption and light conduction. Motion would

deactivate the excited functional states of the pigments and destroy the proper geometric arrangement.

Research on disulphide bonds and thermal stability in T4 lysozyme was presented by R. Wetzel (Biocatalysis Department, Genetech Inc., San Francisco, California). Wild-type T4 lysozyme inactivates by at least four pathways when heated in aqueous solution at pH values near neutrality: (1) reversible unfolding of the polypeptide chain, (2) formation of disulphide-linked oligomers by oxidation of the two cysteine residues, (3) noncovalent aggregation, and (4) nonoxidative covalent changes. Processes (2) and (3) require prior unfolding of the polypeptide, which occurs above the T_m of 63°C; process 4 requires higher temperatures--near 100°C. A disulphide bond engineered into T4 lysozyme between residues 3 and 97, along with replacement of an unpaired cysteine at position 54 stabilizes the molecule against processes (1) to (3). Several lines of evidence show that the stabilizing ability of the disulphide against noncovalent, conformational inactivation (3) is not simply due to its increase of the T_m for reversible folding (1). For example, multiple mutants, containing both the disulphide and a replacement introducing temperature sensitivity, undergo reversible melting at temperatures below wild-type (original organism) but are more stable than wild-type to irreversible inactivation. Preliminary results show that disulphide bonds introduced elsewhere in the molecule produce similar stabilization against irreversible thermal inactivation. The results suggest, according to Wetzel, that, for at least some kinds of stabilization, there may be considerable flexibility in both the global placement of the cross-link and the degree to which the new disulphide must be compatible with local structure.

The topic of denaturation and renaturation of modified lysozymes was discussed by T. Imoto (Faculty of Pharmaceutical Sciences, Kyushu University, Fukuoka, Japan). Lysozyme was variously modified and the stabilities of the derivatives were determined with thermal dena-

turation experiments by Imoto and his group. The contributions of salt bridge, hydrophobic interaction, and cross-linkage were also evaluated. In addition, the stabilities against proteolysis were considered, and Imoto reported that for this stability it was important to depress the rate of unfolding--i.e., to stabilize native conformation. He stated that, as a rule, salt bridges and hydrophobic interactions stabilize native conformation and cross-linkages destabilize denatured conformation. However, cross-linkages are apt to introduce strain in native conformation, and only suitable lengths of cross-linkages can stabilize protein. The stabilization was shown to be generally effective at improving functionality of proteins. Regeneration of disulphide bridges was examined by employing several chemically cross-linked lysozymes. The derivative cross-linked between Lys-13 and Leu-129 showed a 2.3 times accelerated regeneration rate compared with that of unmodified lysozyme. Mutation of Ala-31 to Val completely hindered the regeneration of lysozyme into active conformation. According to Imoto, such a minor mutation can lead to critical damage in the folding of proteins, which is an indispensable process in the genetic engineering of proteins.

Studies on the atomic structure of thymidylate synthetase as a target for rational drug design was presented by R.M. Stroud (Department of Biochemistry and Biophysics, University of California, San Francisco). The atomic structure of thymidylate synthase from *Lactobacillus casei* was determined at 3 Å resolution. The native enzyme is a dimer of identical subunits. The dimer interface is formed by an unusual association between five-stranded beta sheets present in each monomer. Comparison of known sequences with the *Lactobacillus casei* suggested to Stroud and his group that they all have a common core structure around which loops are inserted or deleted in different sequences. Residues from both subunits contribute to each active site. Two arginine side-chains can contribute to binding phosphate on the substrate. It appeared

that the side-chains of several conserved amino acids could account for other determinants of substrate binding.

4 PROTEIN ENGINEERING: METHODS

The use of synthetic oligonucleotides in methods of site-directed mutagenesis, particularly with regard to the needs of protein engineering, was reviewed by M.J. Gait (MRC Laboratory of Molecular Biology, Cambridge, UK). The methods include the use of (1) mismatched oligonucleotide primers to direct single or multiple mutations, deletions, or insertions on single-stranded DNA templates; (2) fully matched oligonucleotide primers in error-forcing repair synthesis; (3) total synthesis of genes; and (4) cassette mutagenesis. The first two methods were used to effect single amino acid changes in the enzyme T4 RNA ligase. For example, substitution of Lys-99 (the site of covalent adenylation in the first step of the enzyme mechanism) by Asn or His leads to inactivity in adenylation as well as in the joining reaction, whereas an Arg-99 mutant retains activity. A total synthesis approach was used by Gait and his group to construct the gene for bovine caltrin (seminal plasmin), a 47 amino acid protein with antifertility and antibacterial activity. The gene has been expressed in *E. coli* as several different fusion proteins with a view to efficient production of caltrin for structural and mutagenesis studies.

The molecular modeling of antibody combining sites was discussed by J.C. Cheetham (Laboratory of Molecular Biophysics, University of Oxford, UK). The lack of detailed structural information on the nature of antibody combining sites, particularly in the complex environment with antigen, has resulted in the widespread application of modeling techniques to the study of these systems. According to Cheetham, an understanding of the nature of the specific interactions between antibody and antigen is fundamentally important if one hopes to design antibodies with a given specificity and affinity for a particular antigen. Such

affinity is perhaps the ultimate goal for such studies, and of immense importance in clinical application.

The modeling of a monoclonal antibody (Mab) called Gloop2 (class IgG), on the basis of DNA sequence information and a structural data base of known immunoglobulin structures, has yielded a preliminary model for the structure of the "anti-loop" binding site. The docking of this model with the loop region of HEL (determined crystallographically at 1.6-Å resolution) and subsequent energy minimization with the GROMOS computer program has given a model for the antibody-antigen complex which may be used--in parallel with site-directed mutagenesis studies of the antibody itself--to probe the nature of the interactions between the two molecules. Prediction of the effects of single and double mutations engineered with the antibody combining site were compared directly with experimental results by Cheetham and her group. The results point to important considerations for the design of any future mutagenesis experiments.

The construction and use of an *E. coli* amber suppressor tRNA bank was presented by J. M. Masson (Molecular Biology Institute, University of California, Los Angeles). A set of amber suppressor genes was constructed for use in amino acid substitution studies as well as protein engineering. These genes were assembled from four to six oligonucleotides, which were annealed and ligated *in vitro* into a plasmid; they are expressed constitutively from a synthetic promoter. Masson and coworkers have constructed high-efficiency suppressors that specifically insert alanine, arginine, cysteine, histidine, and phenylalanine in response to an amber codon. Masson reported that among the additional amber suppressors they have synthesized, some that were inefficient or inserted an unwanted amino acid have been improved in various ways, while others are still being tested for efficiency and specificity of insertion. Masson and his group are examining the effect of a large series of amino acid substitutions in specific proteins by creating amber mutations at known

positions and expressing these mutated genes in the *E. coli* strains that produce the suppressor tRNA's. The lac repressor is one object of the studies by Masson and his group. They have monitored the effects of more than 1000 amino acid substitutions in the repressor by employing all of the existing suppressors, including those constructed in this study.

5 PROTEIN ENGINEERING: BINDING AND CATALYSIS

A review of what has been learned from structure-activity relationships in enzymes was presented by A.R. Fersht (Department of Chemistry, Imperial College of Science and Technology, London, UK). These relationships have been probed now for 5 years by site-directed mutagenesis. Basic information has been derived experimentally on the strengths of intermolecular interactions in solution and their roles in catalysis and molecular recognition. The importance of binding energy in enzyme catalysis has been highlighted by the ability to measure interaction energies between enzymes and substrates throughout the whole reaction profile; i.e., the concept of complementarity has been probed. This in turn has led to a documentation of the types of changes involved in evolution of enzyme catalysis. In particular, the role of hydrogen bonding in these processes has been quantitatively assessed. In addition, protein engineering enables experimental access to measuring inherent properties of proteins such as their dielectric constants.

The effect of active site mutations on mechanism and conformation of Staphylococcal nuclease was presented by J.A. Gerlt (Department of Chemistry and Biochemistry, University of Maryland, College Park). The 1.5-Å x-ray structure of Staphylococcal nuclease suggests that glutamate-43 is appropriately positioned to act as a general base in facilitating the attack of a water molecule on an internucleotide phosphodiester bond of a substrate. Site-directed mutagenesis was used by Gerlt and his group to mutate

residue 43 to aspartate, glutamine, asparagine, serine, and alanine such that the importance of glutamate-43 as a general basic catalyst might be assessed. Although the values for V_{max}/K_m for the mutant enzymes were reduced from 1400-fold (aspartate) to 5000-fold (glutamine, asparagine, serine, and alanine) relative to that measured for the wild-type enzyme, these rate reductions could not be used to quantitate the importance of general basic catalysis. Nuclear magnetic resonance (NMR) spectroscopy revealed that the conformations of the mutant enzymes differ from the conformation of the wild-type enzyme: the chemical shifts and the intensities of the nuclear Overhauser correlations between the aromatic and upfield-shifted methyl resonances were altered by the presence of the point mutation. Gerlt stated that isotopic enrichment of the aromatic and aliphatic amino acids residues with 2H , ^{13}C , and ^{15}N will allow the locations and magnitudes of the conformational changes to be determined. These NMR studies complement the high-resolution x-ray studies of these mutants being carried out in E. Lattman's laboratory at the Johns Hopkins School of Medicine, Baltimore, Maryland.

Studies on the site-directed mutagenesis of Arg-180 of the *crp* gene of *E. coli* K12 were reported by M. C. Serre (Laboratory of Enzymology, CNRS, Gif-sur-Yvette, France). The cyclic AMP receptor protein, CRP (also referred to as a catabolite gene activator protein [CAP]), and its cofactor, cyclic AMP (cAMP), are involved in bacterial regulation, particularly in activating transcription of catabolite operons. A model of interactions between the CRP-cAMP complex and DNA has been proposed. In addition to the established contact between Glu-181 and base pairs 7 and 16 of the CRP recognition site, Serre developed a model that predicts a contact between Arg-180 and base pairs 5 and 18. To verify this prediction, Serre and his group created a set of mutations at codon 180 of the CRP gene of *E. coli* K12, using the oligonucleotide site-directed mutagenesis method. In order to introduce any amino acid at this position, an oligonucleotide was

synthesized in which each of the three positions corresponding to codon 180 was substituted by one of four bases. The resulting mixture served as a mutagenic primer. Mutants were identified by dot-blot hybridization and sequence determination. The yield of mutagenesis was 11 percent, and six classes of mutants were obtained, corresponding to changes of Arg into Cys, Ser, Pro, His, Leu, or Gly. After transfer of the mutated genes on plasmids, these investigators undertook to study the effect of the mutated genes on the expression of genes under CRP control. Preliminary results have confirmed the predicted role of Arg-180 in a specific interaction of CRP with DNA.

Studies of synthetic antibodies with known three-dimensional structure were reported by R. Glockshuber (Gene Center of the University of Munich, Max Planck Institute for Biochemistry, Martinsried [Munich], West Germany). The genes encoding the variable domains (VH and VL) of the phosphorylcholine-binding antibody McPC603, whose crystal structure with and without antigen bound is known, were obtained by DNA synthesis. In addition, Glockshuber and his group constructed genes also encoding the appropriate constant domains of each chain in order to express directly the exact Fab fragment whose crystal structure has been determined. The synthetic genes were designed for facile replacement of gene fragments (e.g., the hypervariable loops) as well as for incorporating current knowledge about efficient expression. These investigators have studied purifications from bacterial expression systems and are comparing their efficiency in obtaining large amounts of protein. The essence of antibody architecture is a framework of fairly constant residues and hypervariable loops or complementarity determining regions that determine specificity--i.e., contain the antigen recognition sequences to a great number of antigens. The particularly well-studied antibody combining site of McPC603 is being used by Glockshuber and his group as a model system for quantitatively investigating factors that contribute to efficient hapten bind-

ing subunit interactions, as well as for the potential of stabilizing a transition state through the controlled modification of the protein. They are attempting to delineate protein contributions to catalysis by making an almost-catalyst (a transition state-binding antibody) perform a suitable hydrolysis reaction.

The topic of protein engineering of yeast triose phosphate isomerase was discussed by G. Pestko (Department of Chemistry, Massachusetts Institute of Technology, Cambridge). The crystal structure of the glycolytic enzyme triose phosphate isomerase from yeast has been determined in high resolution by x-ray crystallography. The structures of complexes of the enzyme with a transition state analogue inhibitor and with its actual substrate have also been determined. Based on these structures a mechanism has been proposed for the catalytic action of these enzymes. Site-directed mutagenesis has been used to alter residues in the active site to test their functions in this mechanism. Mutagenesis has also been employed to test the function of the disordered loop that folds down over the active site when substrates bind. Finally, Pestko and his group have made a series of mutants to investigate the irreversible inactivation of triose phosphate isomerase at high temperatures. Based on the results of these studies, it has been possible to engineer this enzyme to have greatly increased stability to irreversible inactivation at 100°C.

The use of phagemid vectors for expression and mutagenesis of cAMP-dependent protein kinase was reported by M.Z. Zoller (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The phagemid cloning vector consists of hybrid plasmids that contain replication origins from both a bacterial plasmid and a filamentous phage. According to Zoller, these hybrid vectors facilitate DNA sequencing, oligonucleotide-directed mutagenesis, and production of hybridization probes. The vector is usually in a double-stranded form inside the cell. Upon infection of the plasmid-bearing cells with filamentous helper phage, production of single-stranded plasmid molecules is induced.

Single-stranded DNA is prepared as for M13 sequencing and can be used as a template for mutagenesis or sequencing. Zoller and his group have constructed "all-in-one" phagemid vectors for mutagenesis and expression in *E. coli* of the regulatory (R) and catalytic (C) subunits of cAMP-dependent protein kinase (cAdPK) from yeast. Cloning and characterization of the genes for yeast cAdPK was carried out. The expression/mutagenesis vector uses the T7 phage promoter for expression and origin DNA from phage F1. Site-directed mutagenesis was used to probe the interactions between the R and C subunits. Single amino acid changes have been targeted as Ser-145 of the R subunit. This serine residue is thought to play a role in R-C interaction since it is part of the substrate recognition sequence for the C subunit (Arg-Arg-Thr-Ser-Val). Zoller presented data on the use of phagemid vectors for mutagenesis and expression and the specific effects of Ser-145 on R-C interactions.

Studies of protein engineering starting from an immunoglobulin variable domain were presented by H.J. Fritz (Max Planck Institute for Biochemistry, Martinsried). Immunoglobulins are natural precedents of protein engineering: a large spectrum of (binding) properties is realized by the variation of locally confined regions (hypervariable loops) within an otherwise well-conserved and stable structure. Furthermore, the typical immunoglobulin fold is present in a number of seemingly unrelated proteins. The three-dimensional architecture of several immunoglobulins is known, among them dimers of light chain variable domain. Such material was isolated from the urine of patients suffering from Bence-Jones disease. Fritz and his group chose one such variable domain (REI-V) as the starting point of a protein engineering project. A gene for REI-V was chemically synthesized by a combination of known and newly developed techniques. Work currently in progress by Fritz and his group includes: (1) expression of the synthetic gene in bacterial hosts; (2) oligonucleotide-directed mutagenesis of the REI-V gene to obtain structurally predetermined

variants of the protein; and (3) protein-chemical characterization of REI-V and its constructed variants.

A study of the sequence-specific interaction of the cAMP receptor protein (CRP) with DNA was reported by M.E. Gent (Department of Biochemistry and Applied Molecular Biology, University of Manchester, UK). Mutants in the DNA-binding helix F of the CRP were constructed by site-directed mutagenesis. Mutants in a synthetic DNA site derived from the sequence in the *lac* regulatory region were also constructed. The effect of these selected amino acid and nucleotide substitutions on CRP-mediated binding was studied. It was shown that Arg-180 in helix F is crucial for the specific interaction of CRP with DNA whereas Arg-185 and Lys-188 are not involved in specific complex formation. Substitutions of adenine (A) or guanine (G) at position 2 and G at position 5 in the DNA recognition sequence T₁ G₂ T₃ G₄ A₅ both abolish specific binding of a wild-type CRP protein.

A report on evolution guidance for the engineering of alcohol dehydrogenase and ribonuclease was presented by E.G. Weinhold (Laboratory for Organic Chemistry, Institute for Biotechnology Research, Zurich, Switzerland). The evolution of proteins represents (in part) natural selection engineering of a protein to be suitable for different environments, and (in part) the nonfunctional drift of the protein's structure. Thus, some of the information drawn from the recent evolutionary history of a protein, deduced by sequence comparisons of homologous proteins, may assist the biochemist seeking to alter the protein's structure to achieve desired properties, according to Weinhold. This approach was tested by Weinhold and his group on two proteins: yeast alcohol dehydrogenase (ADH) and mammalian ribonuclease (RNase). Several mutants were chosen by comparison of homologous dehydrogenases from yeast and other organisms. A specific example is the change of Arg-211 to Thr. The residue, in a helix remote from the active site, would not be selected for mutagenesis based on a simple inspection of a crystal structure, according to Weinhold.

Yet its alteration was found to affect cofactor binding predictably. In RNase, alterations were introduced into a synthetic gene expressed in *E. coli*. Changes at positions 35 (Leu to Met) and 19 (Ala to Ser) recreate a ribonuclease from a now-extinct organism, the precursor of modern bovovids.

The results of site-directed mutagenesis studies of chloramphenicol acetyltransferase (CAT) were presented by W.V. Shaw (Department of Biochemistry, University of Leicester, UK) in collaboration with A.G.W. Leslie (Department of Biophysics, Imperial College of Science and Technology, London, UK). The many natural variants of CAT in diverse genera of chloramphenicol (CM)-resistant bacteria constitute a family of homologous proteins which inactivate CM by O-acetylation, using coenzyme A as the acyl donor and a ternary complex mechanism with a general base role proposed for the imidazole of His-195. The plasmid-encoded type III enzyme is a stable trimer (3x25 kilodaltons) with identical subunits, each of which contains a conserved acidic residue (Asp-199) which may contribute to catalysis as well as to the stability of CAT enzyme, according to Shaw. Site-directed mutagenesis studies were begun in advance of a structure from crystallographic data with the aim of demonstrating an interaction between His-195 and Asp-199 which might be important for catalysis. Studies of substrate binding, polypeptide folding, and protein stability were also carried out. Shaw thinks that the CAT system may represent a particularly useful vehicle for protein engineering because of the availability of DNA sequences for nine naturally occurring CAT variants and a high-resolution (1.75 Å) structure from x-ray crystallography. The latter locates the CM binding site with great precision and confirms in structural terms the importance of His-195 in the catalytic mechanism proposed by Shaw.

6 PROTEIN ENGINEERING: MEDICAL AND INDUSTRIAL APPLICATIONS

The topic of making therapeutic proteins by directed mutagenesis was re-

viewed by T. Harris (Celltech Ltd., Slough, UK). There are now many classes of proteins which are being synthesized in *E. coli*, yeast or mammalian cells from cloned DNA including hormones, blood proteins, enzymes, inhibitors, and antibodies. Using techniques of site-directed mutagenesis it is possible to change single amino acids in these proteins to make novel molecules with altered characteristics. Some examples are interleukin-1 and α -antiproteinase. The domain structure of many large eukaryotic proteins is often reflected in the genome by the arrangements of the exons and introns making up the gene. As the domains probably fold independently during synthesis it has been possible to change the domain structure of several proteins without destroying specificity of enzymatic activity and conferring novel functions to the hybrid. This is done by manipulating cloned genes or complementary DNA (cDNA) using naturally occurring or introduced restriction sites and oligonucleotides and expressing the chimeras in mammalian cells. Interferons and the plasminogen activators have been altered in this way but antibodies probably provide the best paradigm.

The application of protein engineering in the pharmaceutical industry is now important. Applications of protein engineering to human α_1 -antitrypsin and HIV envelope protein were described by J.M. Lecocq (Transgene SA, Strasbourg, France). Lecocq described two examples which clearly illustrate the high potential of this new technology.

Human α_1 -antitrypsin analogues: the primary function of α_1 -antitrypsin is the inhibition of neutrophil elastase. An α_1 -antitrypsin deficiency results in lung emphysema. Moreover, the α_1 -antitrypsin from lungs of heavy smokers is inactivated due to the oxidation of the methionine (Met) residues at the active site. A (Met-358 to Val-358) α_1 -antitrypsin analogue which remains fully active as an elastase inhibitor but which is also resistant to oxidative inactivation was constructed by Lecocq and coworkers and characterized. The inhibition of the clotting pathway by other analogues with

modified residues in the active site has also been studied.

HIV envelope protein: the envelope glycoprotein represents the major antigen at the surface of HIV viral particles, and efforts to produce a vaccine against AIDS have centered on this protein. Lecocq has constructed a recombinant vaccinia virus expressing the envelope protein. In infected tissue culture cells the envelope precursor, gp160, is efficiently cleaved to two components, an extracellular one (gp120) and a transmembrane segment (gp41). Lecocq has observed that gp120 is rapidly shed from the cell surface, thus offering an explanation for the low immunogenicity of the envelope protein. In order to enhance the immunogenicity of the envelope protein, Lecocq and his group have genetically engineered new variants which are now being checked.

Studies of a cassette gene for tissue plasminogen activator (t-PA) as a model for investigating protein domain function were presented by E.F. Rehberg (The Upjohn Company, Kalamazoo, Michigan). Tissue plasminogen activator (t-PA) is a serine protease that converts the inactive zymogen plasminogen to plasmin, the major fibrin-degrading enzyme in the fibrinolytic system. Tissue t-PA is a multidomain protein consisting of five putative domains: the fibronectin finger domain (F), the epidermal growth factor domain (G), the kringle 1 domain (K), the kringle 2 domain (K), and the protease domain (P). Rehberg and his group have chemically synthesized a portion of the t-PA gene to allow the incorporation of convenient restriction sites between each of the domains, thus simplifying their manipulation. The system not only provides for the construction and expression of t-PA mutants having single domain deletions but also facilitates the site-directed mutagenesis and expression of individual domains. By using this approach, these investigators have constructed and expressed the prototype, full-length molecule (GFK₁K₂P) as well as the deletion mutants FK₁K₂P, GFK₂P, GFK₁P, and FK₂P. In addition, a number of mutants with multiple domain deletions, additions and transpositions have

been made. Rehberg and his group have purified each of these mutants on immunoaffinity columns and have assessed molecular weight by fibrin autoradiography, amidolytic activity in the presence and absence of CNBr-digested fibrinogen fragments, and specific activity via (1,3-³H) di-isopropyl phosphofluoride active site titration. Their analyses showed that most gross domain deletions do not appear to have dramatic effects on specific enzyme activity. According to Rehberg, these data suggest further that the interdomain amino acid modifications made as a result of inserting unique restriction sites do not appear to have any deleterious effect on the biochemical properties of t-PA examined thus far.

The production of monomeric insulins by protein engineering was reported by J. Brange (Novo Research Institute, Bagsvaerd, Denmark). Insulin associates above physiological concentrations of 10⁻¹⁰ M into dimers and hexamers. This strong tendency to self-assembly has numerous advantages in relation to the events following biosynthesis in the beta cell of the pancreas, but does not necessarily represent useful properties for the therapeutic formulation and clinical use of insulin, according to Brange. Therefore, Brange and his group applied protein engineering to the hormone with the aim of counteracting self-association and changing its physicochemical properties. Substitutions were performed at five different residues (B9, B12, B26, B27, and B28) by oligonucleotide-directed mutagenesis or by total gene synthesis. The insulin analogues were produced by fermentation of single chain precursors in yeast followed by semisynthetic conversion into the respective mutants. As assessed by gel filtration, measurements of osmotic pressure and circular dichroism (CD) spectroscopy, mutation of B12-Val → Ile, B26-Tyr → Glu, or B28-Pro → Glu resulted in analogues which remained essentially monomeric even above millimolar concentrations. Molecular modeling has revealed that electrostatic effects probably are the main explanation for the observed prevention of the association, but steric hindrance also accounts in some cases for

the weakened self-assembly. Thus, monomeric insulins can be obtained by a single mutation in the monomer-monomer interface of the insulin dimer.

Diphtheria toxin receptor binding domain substitution with interleukin-2 was reported by D. Williams (Seragen Inc., Hopkinton, Massachusetts). The mature form of diphtheria toxin (molecular weight 58,348) is cleaved by trypsin into fragment A (mol wt 21,167) and fragment B (mol wt 37,199). The catalytic center for ADP-ribosyl transferase activity is carried by fragment A, while the function of fragment B is to bind the toxin molecule to the cell surface and facilitate the membrane translocation of fragment A into the cytosol. Receptor binding domain substitution should result in the formation of new recombinant toxin molecules whose target specificity is determined by the ligand component of the chimeric protein, according to Williams. Thus, he and his group genetically fused a cDNA insert encoding amino acids 2 to 133 of interleukin-2 (IL2) to amino acid 465 of diphtheria toxin. The resulting chimeric toxin has a molecular weight of 68,086 and has antigenic determinants to both diphtheria toxin and IL2. The toxin-related IL2 fusion protein (IL2-toxin) was expressed in recombinant *E. coli* K12.

Following purification from the periplasmic compartment, IL2-toxin is selectively cytotoxic for eukaryotic cells which bear high-affinity IL2 receptors. Cytotoxicity is blocked by excess IL2, monoclonal antibodies to the IL2 receptor, and lysozomotropic reagents.

7 CONCLUSION

This informative and intensive conference on the new and rapidly developing area of protein engineering covered many of the protein-based aspects of this subject as well as some of the sophisticated molecular biological techniques needed to make and express mutant proteins. In addition, the concepts of making defined mutations in proteins to probe ligand binding and protein structure/function relationships was also covered. It is evident that the prospect for applications of recombinant DNA techniques in the study and use of enzyme catalysis and protein-protein interactions has been radically changed in the last 4 years. Protein engineering research has also become an area of top priority for industrial organizations--in particular, the pharmaceutical industry. Some applications to commercial products were also presented by the speakers.

END

9-87

DTIC